

Simple optimality proofs for Least Recently Used in the presence of locality of reference

Citation for published version (APA):

Hiller, B., & Vredevelde, T. (2009). *Simple optimality proofs for Least Recently Used in the presence of locality of reference*. METEOR, Maastricht University School of Business and Economics. METEOR Research Memorandum No. 055 <https://doi.org/10.26481/umamet.2009055>

Document status and date:

Published: 01/01/2009

DOI:

[10.26481/umamet.2009055](https://doi.org/10.26481/umamet.2009055)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Benjamin Hiller, Tjark Vredevelde

**Simple optimality proofs for
Least Recently Used in the
presence of locality of reference**

RM/09/055

METEOR

Maastricht University School of Business and Economics
Maastricht Research School of Economics
of Technology and Organization

P.O. Box 616
NL - 6200 MD Maastricht
The Netherlands

Simple optimality proofs for Least Recently Used in the presence of locality of reference

Benjamin Hiller* Tjark Vredeveld**

November 5, 2009

Abstract

It is well known that competitive analysis yields results that do not reflect the observed performance of online paging algorithms. Many deterministic paging algorithms achieve the same competitive ratio, ranging from inefficient strategies as flush-when-full to the well-performing least-recently-used (LRU).

In this paper, we study this fundamental online problem from the viewpoint of stochastic dominance. We give simple proofs that when sequences are drawn from distributions modelling locality of reference, LRU stochastically dominates any other online paging algorithm. As a byproduct, we obtain simple proofs of some earlier results.

1 Introduction

The *paging* problem is one of the most fundamental problems in online optimization. It models an optimization problem occurring in a two-level memory system. The first level, called the *slow memory*, stores a fixed set M of pages and the second level, which is the *fast memory* or *cache*, contains up to k pages of the set M . We will also refer to k as the *cache size*. In paging, one needs to serve a sequence of requests for pages $\sigma \in M^n$, where n is the number of pages requested. To serve a request for page $p \in M$, the system needs to have this page in the cache. If a requested page is not in the cache a *page fault* occurs. The requested page must then be loaded into the cache, and whenever the cache contains k pages, at least one page must be evicted from it. A paging algorithm needs to decide which page(s) will be evicted from the cache on a page fault. The goal is to minimize the number of page faults. Standard paging strategies include the following.

- *Least recently used* (LRU): whenever there is a page fault, evict the page whose most recent request was earliest.
- *First in first out* (FIFO): on a page fault, evict the page that has been in the cache the longest.
- *Flush when full* (FWF): On a page fault, when the cache is full, it evicts all pages from the cache.

*Zuse Institute Berlin, Takustraße 7, D-14195 Berlin, Germany, hiller@zib.de

**Maastricht University, Department of Quantitative Economics, P.O.Box 616, 6200 MD, Maastricht, The Netherlands, t.vredeveld@maastrichtuniversity.nl

- *Longest forward distance* (LFD): On a page fault, LFD evicts the page whose next request is farthest in the future.

All these algorithms, except LFD, are online algorithms. That is, they decide which page to evict without knowing future requests or the number of requests in the sequence. As LFD needs to know the future, it cannot be used in practice. However, LFD is an optimal page eviction strategy [6]. Both LRU as well as FWF are *marking algorithms*. Marking algorithms label a page ‘marked’ whenever it is requested. On a page fault and when the cache contains k pages, a marking algorithm evicts an unmarked page. As soon as the cache contains k marked pages, a marking algorithm labels all pages as unmarked. Note that FIFO is not a marking algorithm.

The standard yardstick for online algorithms has become *competitive analysis* [23, 18]. An online algorithm is called *c-competitive* if, for all request sequences, the cost of the algorithm, i. e., the number of page faults is at most c times the optimal offline costs. The smallest c for which an algorithm is c -competitive is also known as the *competitive ratio*. Sleator and Tarjan [23] showed that LRU and FIFO have a competitive ratio of k and that this is the best possible. Karlin et al. [18] gave a different proof for the same results and in addition they showed that FWF also has a competitive ratio of k . Torng [24] extended these results showing that all deterministic marking algorithms are k -competitive.

Related work. As FIFO, LRU and all other marking algorithms have the same competitive ratio, competitive analysis obviously fails to distinguish between these algorithms although they empirically perform very differently. Therefore, there has been research on the refinement of competitive analysis and alternative models for assessing online paging algorithms. Young [25, 27] introduced the notion of *loose competitiveness*, in which paging algorithms are evaluated for varying sizes of the cache, ignoring input sizes that have a high competitive ratio only for few cache sizes. He showed that several deterministic paging strategies are loosely $\mathcal{O}(\log k)$ competitive. The Max/Max ratio of Ben-David and Borodin [7] compares the worst case amortized behavior of an algorithm with that of an optimal offline algorithm. An algorithm is said to have Max/Max ratio c if it is guaranteed that on no request sequence will it ever have to pay more than c times the maximal cost that an optimal offline algorithm pays on a sequence of the same length. Koutsoupias and Papadimitriou [20] introduced the diffuse adversary. In the concept of diffuse adversary, an average case competitive analysis is performed, but instead of selecting any probability distribution on the input sequence the diffuse adversary may only select a probability distribution from a prespecified class of distributions. Young [26] also performed a diffuse adversary analysis for the paging problem, showing the optimality of LRU against some specific diffuse adversary. In the relative worst-order ratio [10] two algorithms are compared each on their respective worst-case permutation of a request sequence. Boyar et al. [10] showed that LRU is better than FWF, but LRU and FIFO are equally good according to this measure. Recently, bijective analysis was introduced [3, 4]. In bijective analysis, one tries to find a bijective mapping from the set of instances on itself such that the preferred algorithm delivers on each instance a better objective function value than the algorithm, to which it is compared, has on the mapped instance. Angelopoulos and Schweitzer [4] showed that LRU is an optimal algorithm under

this framework for a restricted class of sequences.

One of the reasons that competitive analysis is not able to make a distinction between the performance of several paging algorithms is that it considers arbitrary request sequences. In practice, however, request sequences have some structure, for instance they often feature *locality of reference*. In a strict sense this means that if a page is referenced, it is likely to be referenced again in the near future, a fact that has been observed early and formalized by Denning with his working set concept [12, 13]. Based on this concept, Torng [24] introduces a model for locality of reference by lower bounding the length of a subsequence containing a certain number of different pages. He shows that, among other algorithms, LRU achieves a constant competitive ratio. Albers et al. [2] gave another model for locality of reference, also based on Denning’s working set concept. They showed that LRU is an optimal online algorithm in their model and that FIFO and marking strategies are not optimal in general. Becchetti [5] performs a diffuse adversary analysis, where the diffuse adversary is only allowed to choose a probability distribution favoring recently over less recently requested pages. He shows that in this model, LRU outperforms FWF. In fact, this model was known as LRU-stack model for a long time [11].

There is also the broader sense of locality of reference, meaning that usually “each time a page is referenced by a program, the next page to be referenced is very likely to come from a small set of pages” [9]. To formalize this, Borodin et al. [9] presented the *access graph model*, in which a graph models which pages can be requested after a certain page has been asked. Using this model, they showed that LRU is at least as good as FIFO. This approach was extended in several ways in [17, 19, 15]. Panagiotou and Souza [22] introduce another model, restricting the sequences such that successive references to the same page are mostly close together or far apart.

Our results. One of the weaknesses of competitive analysis is that it fails to distinguish between all kinds of algorithms. Therefore, alternative measures for the performance of online algorithms are needed. In this paper, we compare the performance of paging algorithms on random input sequences directly using stochastic dominance. This method for comparing online algorithms has been introduced in [16]. Given a probability distribution on all possible input sequences, we let X^{Alg} denote the random variable of the number of pages faults of an online algorithm Alg. We say that online algorithm Alg_1 is stochastically better than online algorithm Alg_2 if the random variable X^{Alg_1} is stochastically dominated by X^{Alg_2} , i. e., $\Pr[X^{\text{Alg}_1} \geq x] \leq \Pr[X^{\text{Alg}_2} \geq x]$ for all $x \in \mathbb{R}$.

We study two kinds of distribution functions on the input sequences which both model locality of reference in the strict sense. The first kind of probability distributions are special LRU-stack model distributions also considered by Becchetti [5], which give higher probability to pages recently asked than to those asked further in the past. The second kind of distributions are uniform distributions on input sequences that fulfill a very broad definition of locality of reference. This class includes the f -consistent sequences according to Albers et al. [2] and a generalization of the model by Torng [24] used in [14]. For both kinds of distributions we give simple proofs that LRU is stochastically better than any other online algorithm. As a byproduct, we obtain simple proofs for various results known before, among them the bijective analysis result in by

Angelopoulos and Schweitzer [4].

2 Mode of analysis and locality of reference models

We start by recalling some basic notions for paging algorithms. A standard tool is the partitioning of a sequence into *phases*, see e. g., [8]. The first phase starts with the first request. Phase ℓ starts with the $(k + 1)$ st *distinct* request after the start of phase $\ell - 1$. Each phase ends just before the start of the next phase or at the end of the sequence, whichever comes first.

Given a request sequence σ , we say that a page p is *marked w. r. t.* σ if it has been requested in the final phase of σ ; otherwise, we say that p is *unmarked w. r. t.* σ . Note that by definition of the phases, there cannot be more than k marked pages w. r. t. a request sequence. Also note that all pages are unmarked w. r. t. the empty sequence. Moreover, the partition into phases and the set of marked pages at any point in the sequence do not depend on the algorithm. Observe that a marking algorithm has, at any point in time, all marked pages in its cache, which justifies the name.

A paging algorithm is called *lazy* if it evicts a page only on a page fault and never evicts more than one page.

2.1 Stochastic dominance analysis of online algorithms

The competitive ratio as a measure of the performance of an online algorithm has been criticized for failing to discriminate between algorithms that perform very differently in practice.

In our approach [16] we compare the performance of algorithms on random request sequences drawn according to certain probability distributions. In contrast to competitive analysis or diffuse adversary analysis, we directly compare two algorithms to each other without referring to an optimal offline solution. We compare the performance of online algorithms using *stochastic dominance*, a well-known stochastic order. A random variable X is said to be *stochastically dominated* by a random variable Y , written $X \leq_{\text{st}} Y$, if

$$\Pr[X \geq x] \leq \Pr[Y \geq x] \quad \text{for all } x \in \mathbb{R}. \quad (1)$$

One way to think of this approach is that we compare the *distributions* of the performances of two online algorithms instead of aggregate statistics like the expected value or the maximum. We will later see that in some cases, there are distribution-free interpretations of a stochastic dominance result.

Stochastic dominance has some interesting properties [21]. Abusing notation, we denote the random variable for the performance of an algorithm Alg by the same symbol, Alg. The first interesting consequence of stochastic domination of the performance of one algorithm by another, i. e., $\text{Alg}_1 \leq_{\text{st}} \text{Alg}_2$, is that the expected performance of Alg_1 is also better than that of Alg_2 , i. e., $\mathbb{E}[\text{Alg}_1] \leq \mathbb{E}[\text{Alg}_2]$. This again implies that the average competitive ratio of Alg_1 is not worse than that of Alg_2 : $\mathbb{E}[\text{Alg}_1]/\mathbb{E}[\text{Opt}] \leq \mathbb{E}[\text{Alg}_2]/\mathbb{E}[\text{Opt}]$. Finally, if we have an non-decreasing function f on the possible outcomes for Alg_1 and Alg_2 , then $f(\text{Alg}_1) \leq_{\text{st}} f(\text{Alg}_2)$. This can be used to conclude that

Alg_1 is also better than Alg_2 in the full access cost model [24], where a page in the cache incurs cost 1, when requested, and a page fault incurs cost $1 + p$ for some parameter $p > 0$.

In case $\text{Alg}_1 \leq_{\text{st}} \text{Alg}_2$ holds for the uniform distribution on (a subset of) the sequences, this is equivalent to the existence of a bijective mapping ϕ on the sequences such that $\text{Alg}_1(\sigma) \leq \text{Alg}_2(\phi(\sigma))$ for any σ . This strong way of comparing online algorithms is called *bijective analysis* and has been introduced in [3]. Thus stochastic dominance results for the uniform distribution share the favorable properties of bijective analysis results discussed in [3].

2.2 Paging with locality of reference

We consider the three models for paging with locality of reference in the strict sense that were mentioned in the introduction [24, 11, 2]. The main result of this paper is that LRU is optimal w.r.t. stochastic dominance for sequences exhibiting this kind of locality of reference. It is well known that LRU is not an optimal online algorithm for sequences with locality of reference in the broader sense, since e.g., the access graph model allows the $k + 1$ cycle as an access graph.

The age model Coffman and Denning [11] introduced the following probabilistic model which we call the *age model*. In the age model, the next request for a prefix sequence σ is generated based on the age of the pages. For a prefix sequence σ , the age of a page $p \in M$ is defined by

$$\text{age}(p, \sigma) := \begin{cases} l & \text{if } p \text{ is the } l\text{th most recently requested page,} \\ \infty & \text{if } p \text{ does not appear in } \sigma. \end{cases}$$

We say that a probability distribution over the request sequences is an *age model distribution* if it arises in the following way. Let Δ be the set of probability distributions over $\{1, \dots, |M|\}$. Given a prefix sequence σ , the probability $\Pr[p | \sigma]$ is determined by an *age distribution* $\delta \in \Delta$. The age distribution δ gives the age of the new request page p , i.e., if a is a realization according to δ , the next page is $p \in M$ with $\text{age}(p, \sigma) = a$. If there is no page with age a one of the pages with age ∞ is chosen arbitrarily.

Let $\mathcal{D} \subseteq \Delta$ be the set of distributions with monotone non-increasing distribution functions. Note that considering age distributions from \mathcal{D} models locality of reference: Pages requested more recently have a high probability to be requested next. Note that \mathcal{D} contains two out of three classes of age distributions studied by Becchetti [5].

The concave function model and the a -locality model In contrast to the age model, the concave function model [2] and the a -locality model [24] are deterministic models which restrict the set of request sequences.

Albers et al. [2] propose the *concave function model* which models working sets. Locality of reference is modeled by an increasing concave function $f: \mathbb{N} \rightarrow \mathbb{R}$, which specifies the maximum number $f(l)$ of distinct pages in a (contiguous) subsequence of length l for any $l \in \mathbb{N}$. A request sequence for which each subsequence of length l has at most $f(l)$ distinct pages is called *f -consistent*.

Torng [24] generalizes the partitioning of a sequence to m -phases, i.e., a phase consists of m distinct pages. For any sequence σ , define the phase indices $I_{i,m}(\sigma)$ recursively by

$$I_{0,m}(\sigma) = 0, \\ I_{i+1,m}(\sigma) = \max\{j \leq n \mid \sigma[I_{i,m}(\sigma), j] \text{ contains at most } m \text{ distinct pages}\}.$$

The m -phases of σ are then given by $\sigma[I_{0,m}(\sigma)+1, I_1(\sigma)], \sigma[I_{1,m}(\sigma)+1, I_{2,m}(\sigma)], \dots$, where $\sigma[i, j] = \sigma_i, \dots, \sigma_j$ denotes the corresponding subsequence of σ . Denote by $A(\sigma, m)$ the average length of the m -phases in σ . Torng argues that σ exhibits significant locality of reference if $A(\sigma, m) \gg m$. To capture this formally, we introduce the notion of a -locality for a function $a: \mathbb{N} \rightarrow \mathbb{R} \in \Omega(1)$. A sequence σ is called a -local if $A(\sigma, m) \geq a(m)m$ holds for all $m = 1, \dots, |\sigma|$. Note that this generalizes the notion of “ a -local” used in [14].

3 Optimality of LRU for paging with locality of reference

In this section we show that LRU is optimal w.r.t. to stochastic dominance, i.e., incurs stochastically fewer page faults than any other paging algorithm, for a certain class of request sequence distributions. This class includes all distributions according to the age model as well as the uniform distribution over all f -consistent sequences for any increasing concave function f and the uniform distribution over all a -local sequences for any function $a: \mathbb{N} \rightarrow \mathbb{R} \in \Omega(1)$.

3.1 Preliminaries

For an online algorithm Alg, an integer $j = 1, \dots, n$ and a sequence σ of length $|\sigma| = j - 1$, the random variable $X_j^{\text{Alg}}(\sigma) = 1$ if the first $j - 1$ requests are given by σ and Alg encounters a page fault on the j th request; otherwise $X_j^{\text{Alg}}(\sigma) = 0$. The random variable $Y_j(\sigma) = 1$ if $|\sigma| = j$ and the first j requests are as in σ ; otherwise $Y_j(\sigma) = 0$. Moreover, we define the random variable $W^{\text{Alg}}(\sigma) = (X_j^{\text{Alg}}(\sigma) \mid Y_{j-1}(\sigma) = 1)$, i.e., $W^{\text{Alg}}(\sigma) = 1$ if the next request after processing $j - 1$ requests from σ leads to a page fault in Alg, and $W^{\text{Alg}}(\sigma) = 0$ otherwise.

Given a sequence σ of length $|\sigma| \geq j$ and an online algorithm Alg, the variable $Z_j^{\text{Alg}}(\sigma) = 1$ if the j th request leads to a page fault when Alg operates on σ and $Z_j^{\text{Alg}}(\sigma) = 0$ otherwise. Note that $Z_j^{\text{Alg}}(\sigma)$ is deterministically determined by Alg, j , and σ . On a sequence $\sigma \in M^n$, an online algorithm Alg has $\sum_{j=1}^n Z_j^{\text{Alg}}(\sigma)$ page faults. Therefore, the random variable Alg giving the total number of page faults on a sequence of length n can be written as

$$\text{Alg} = \sum_{\sigma \in M^n} \sum_{j=1}^n Z_j^{\text{Alg}}(\sigma) Y_n(\sigma).$$

Finally, we denote by $C^{\text{Alg}}(\sigma)$ the set of pages in the cache after the sequence σ has been processed by algorithm Alg.

Lemma 3.1 *The value of an online algorithm Alg can be written as*

$$\text{Alg} = \sum_{j=1}^n \sum_{\sigma \in M^{j-1}} X_j^{\text{Alg}}(\sigma).$$

Proof.

$$\begin{aligned} \text{Alg} &= \sum_{\sigma \in M^n} \sum_{j=1}^n Z_j^{\text{Alg}}(\sigma) Y_n(\sigma) = \sum_{j=1}^n \sum_{\sigma \in M^n} Z_j^{\text{Alg}}(\sigma) Y_n(\sigma) \\ &= \sum_{j=1}^n \sum_{\sigma_1 \in M^{j-1}} \sum_{p \in M} \sum_{\sigma_2 \in M^{n-j}} Z_j^{\text{Alg}}(\sigma_1 p \sigma_2) Y_n(\sigma_1 p \sigma_2) \end{aligned}$$

as Alg is an online algorithm, we have

$$= \sum_{j=1}^n \sum_{\sigma_1 \in M^{j-1}} \sum_{p \in M} Z_j^{\text{Alg}}(\sigma_1 p) \sum_{\sigma_2 \in M^{n-j}} Y_n(\sigma_1 p \sigma_2)$$

due to the fact that $Y_j(\sigma_1) = \sum_{\sigma_2 \in M^{n-j}} Y_n(\sigma_1 \sigma_2)$, we can write

$$\begin{aligned} &= \sum_{j=1}^n \sum_{\sigma_1 \in M^{j-1}} \sum_{p \in M} Z_j^{\text{Alg}}(\sigma_1 p) Y_j(\sigma_1 p) \\ &= \sum_{j=1}^n \sum_{\sigma_1 \in M^{j-1}} X_j^{\text{Alg}}(\sigma_1), \end{aligned}$$

where the last equality follows from that fact that $X_j^{\text{Alg}}(\sigma_1) = 1$ for all realizations of the request sequence that start with σ_1 and the j th request p leads to a page fault, i. e., $Z_j^{\text{Alg}}(\sigma_1 p) = 1$. \square

The following theorem is the main result underlying our stochastic dominance proofs. We call a probability distribution over request sequences a *prefix distribution* if they are completely described by the probability that page p is requested given that the sequence up to this page is σ , $\Pr[p | \sigma]$. Note that all age model distributions are of this type.

Theorem 3.2 *Let Alg₁ and Alg₂ be two online paging algorithms. Assume that the request sequence is drawn according to a prefix distribution and denote by P the random next request after prefix σ . Suppose that*

$$\Pr[P \in C^{\text{Alg}_1}(\sigma)] \geq \Pr[P \in C^{\text{Alg}_2}(\sigma)] \quad (2)$$

for any sequence σ . Then $\text{Alg}_1 \leq_{\text{st}} \text{Alg}_2$.

Proof. By Lemma 3.1, it is sufficient to show

$$X_j^{\text{Alg}_1}(\sigma) \leq_{\text{st}} X_j^{\text{Alg}_2}(\sigma),$$

for all $j = 1, \dots, n$, and all sequences $\sigma \in M^{j-1}$. As the variables $X_j^{\text{Alg}}(\sigma)$ are binary random variables, this is equivalent to

$$\Pr[X_j^{\text{Alg}_1}(\sigma) = 1] \leq \Pr[X_j^{\text{Alg}_2}(\sigma) = 1]. \quad (3)$$

For any online algorithm Alg, $j \in \{1, \dots, n\}$, and $\sigma \in M^{j-1}$, we can write

$$\begin{aligned} \Pr[X_j^{\text{Alg}}(\sigma) = 1] &= \Pr[Y_{j-1}(\sigma) = 1 \wedge X_j^{\text{Alg}}(\sigma) = 1] \\ &= \Pr[X_j^{\text{Alg}}(\sigma) = 1 \mid Y_{j-1}(\sigma) = 1] \cdot \Pr[Y_{j-1}(\sigma) = 1] \\ &= \Pr[W^{\text{Alg}}(\sigma) = 1] \cdot \Pr[Y_{j-1}(\sigma) = 1]. \end{aligned}$$

Since $Y_{j-1}(\sigma)$ does not depend on the algorithm, (3) is equivalent to $\Pr[W^{\text{Alg}_1}(\sigma)] \leq \Pr[W^{\text{Alg}_2}(\sigma)]$. Moreover, $\Pr[W^{\text{Alg}_1}(\sigma)] \leq \Pr[W^{\text{Alg}_2}(\sigma)]$ is equivalent to (2). \square

3.2 Optimality results

Theorem 3.3 *Suppose the request sequence is chosen according to the age model with age distributions from \mathcal{D} . Then the number of page faults of LRU is stochastically dominated by that of any online algorithm.*

Proof. Let Alg be any online paging algorithm. We show that condition (2) of Theorem 3.2 is satisfied. Note that always $|C^{\text{LRU}}(\sigma)| \geq |C^{\text{Alg}}(\sigma)|$ holds. By definition of LRU, there is an injective mapping $\phi: C^{\text{Alg}}(\sigma) \rightarrow C^{\text{LRU}}(\sigma)$ that maps a page from $C^{\text{Alg}}(\sigma)$ to a page in $C^{\text{LRU}}(\sigma)$ which is not older. Let p be some page that Alg has in the cache and denote by P the random next request generated according to the age model distribution. Clearly, $\Pr[P = \phi(p)] \geq \Pr[P = p]$, which implies condition (2). \square

Remark *i.)* Coffman and Denning [11, p. 276] show that LRU achieves an optimal expected number of page faults if the request sequence is generated as in Theorem 3.3. This result is implied by Theorem 3.3. *ii.)* The result actually holds for all age distributions $(p_1, \dots, p_{|M|})$ with $\min\{p_1, \dots, p_k\} \geq \max\{p_{k+1}, \dots, p_{|M|}\}$. As the uniform distribution over all sequences belongs to the family of age model distributions, we also have that *iii.)* LRU, as any other lazy paging algorithm, is an optimal algorithm w.r.t. bijective analysis for the set of all request sequences.

We will now prove similar optimality results for the deterministic locality of reference models mentioned in Section 2.2, assuming the uniform distribution over the feasible sequences. We start by explaining how we can use a prefix distribution to generate a uniform distribution. Denote by S_n the set of sequences of length n that are feasible for the locality of reference model (i.e., f -consistent or a -local). Moreover, for any σ , $0 \leq |\sigma| < n$, and $p \in M$ let $S_n(\sigma, p)$ be the set of extensions of σp to a feasible sequence of length n , i.e.,

$$S_n(\sigma, p) = \{\sigma' \in M^{n-|\sigma|} \mid \sigma p \sigma' \in S_n\}.$$

Finally define $L_n(\sigma, p) = |S_n(\sigma, p)|$ and $L_n(\sigma) = \sum_{p \in M} L_n(\sigma, p)$. A uniformly distributed random sequence $\Sigma \in S_n$ may be generated as follows: after $\sigma = (\sigma_1, \dots, \sigma_{i-1})$ has been chosen, we set $\sigma_i = p$ with probability $L_n(\sigma, p)/L_n(\sigma)$ for each $p \in M$.

Lemma 3.4 *Consider any online paging algorithm Alg and suppose that for any sequence σ , $0 \leq |\sigma| < n$, and for any pages $p \in C^{\text{LRU}}(\sigma) \setminus C^{\text{Alg}}(\sigma) =: P$ and $q \in C^{\text{Alg}}(\sigma) \setminus C^{\text{LRU}}(\sigma) =: Q$ we have*

$$L_n(\sigma, p) \geq L_n(\sigma, q). \quad (4)$$

Then $\text{LRU}(\Sigma) \leq_{\text{st}} \text{Alg}(\Sigma)$, where Σ is a random sequence distributed uniformly over S_n .

Proof. For a sequence σ with $|\sigma| = j < n$ let P denote the random request that follows σ according to the distribution defined above, i. e., $\Pr[P = p] = L_n(\sigma, p)/L_n(\sigma)$. By Theorem 3.2, it is sufficient to show $\Pr[P \in C^{\text{LRU}}(\sigma)] \geq \Pr[P \in C^{\text{Alg}}(\sigma)]$ for any sequence σ . We have

$$\begin{aligned} \Pr[P \in C^{\text{LRU}}(\sigma)] &= \frac{\sum_{p \in C^{\text{LRU}}(\sigma)} L_n(\sigma, p)}{L_n(\sigma)} \\ &= \frac{\sum_{p \in C^{\text{LRU}}(\sigma) \cap C^{\text{Alg}}(\sigma)} L_n(\sigma, p) + \sum_{p \in P} L_n(\sigma, p)}{L_n(\sigma)} \\ &\geq \frac{\sum_{p \in C^{\text{LRU}}(\sigma) \cap C^{\text{Alg}}(\sigma)} L_n(\sigma, p) + \sum_{q \in Q} L_n(\sigma, q)}{L_n(\sigma)} \\ &= \Pr[P \in C^{\text{Alg}}(\sigma)], \end{aligned}$$

where we used (4) and the fact that $|C^{\text{LRU}}(\sigma)| \geq |C^{\text{Alg}}(\sigma)|$ which implies $|P| \geq |Q|$. \square

In the following proofs we need the following notation borrowed from [4]. For a sequence σ and two distinctive pages $p, q \in M$, the *complement sequence* of σ w. r. t. p and q , denoted by $\bar{\sigma}^{(p,q)}$, arises from σ by exchanging p and q and keeping all other requests. Note that the mapping $\sigma \mapsto \bar{\sigma}^{(p,q)}$ is bijective since it is self-inverse.

Theorem 3.5 *Let $f: \mathbb{N} \rightarrow \mathbb{R}$ be an increasing concave function and let Σ be a request sequence drawn uniformly at random from all f -consistent sequences of length n . Then $\text{LRU}(\Sigma) \leq_{\text{st}} \text{Alg}(\Sigma)$.*

Proof. We apply Lemma 3.4 and show $L_n(\sigma, p) \geq L_n(\sigma, q)$ for any sequence σ , $0 \leq |\sigma| < n$ and for any pages $p \in P$ and $q \in Q$ by giving an injective map from $S_n(\sigma, q)$ to $S_n(\sigma, p)$. In particular, we claim that for $\sigma' \in S_n(\sigma, q)$, $\sigma' \mapsto \bar{\sigma}'^{(p,q)}$ is such a map.

By definition of p and q , the last request for page p in σ was after the last request for page q . Let $\omega = \sigma q \sigma'$ be the f -consistent extended sequence corresponding to σ' . Lemma 1 from [4] states that either the sequence $\sigma p \bar{\sigma}'^{(p,q)}$ is f -consistent or the first access to p after σ in ω is before the first access to q . Since in ω , q is the first requested page after σ , the second case obviously cannot be true and therefore $\bar{\sigma}'^{(p,q)}$ must be an element of $S_n(\sigma, p)$. \square

Theorem 3.6 *Consider a function $a: \mathbb{N} \rightarrow \mathbb{R} \in \Omega(1)$ and let Σ be a request sequence drawn uniformly at random from all a -local sequences of length n . Then $\text{LRU}(\Sigma) \leq_{\text{st}} \text{Alg}(\Sigma)$.*

Proof. We use the same proof technique as in the last theorem, using again the mapping $\sigma' \mapsto \bar{\sigma}'^{(p,q)}$ for each $\sigma' \in S_n(\sigma, q)$. Let $\omega := \sigma q \sigma'$ be the original sequence and $\bar{\omega} := \sigma p \bar{\sigma}'^{(p,q)}$ be its mapping image. We establish $A(\bar{\omega}, m) \geq A(\omega, m)$ for any m by showing $I_{i,m}(\bar{\omega}) \geq I_{i,m}(\omega)$ by induction on i . Let ℓ be the number of m -phases of σ . It is obvious that $I_{i,m}(\bar{\omega}) = I_{i,m}(\omega)$ for all $i < \ell$.

Suppose $|\sigma| = l$ and denote by $D(\sigma, i, j)$ the set of distinct pages in the subsequence $\sigma_i, \dots, \sigma_j$. We claim $I_{\ell, m}(\bar{\omega}) \geq I_{\ell, m}(\omega)$, or equivalently,

$$|D(\bar{\omega}, I_{\ell-1, m}(\sigma) + 1, I_{\ell, m}(\omega))| \leq |D(\omega, I_{\ell-1, m}(\sigma) + 1, I_{\ell, m}(\omega))|. \quad (5)$$

If q is referenced in $\sigma[I_{\ell-1, m} + 1, l]$, then so is p since $p \in C^{\text{LRU}}(\sigma)$ and $q \notin C^{\text{LRU}}(\sigma)$. This implies that (5) holds with equality in this case. The same is true if both p and q do not occur in $\sigma[I_{\ell-1, m} + 1, l]$. In the case that p occurs in $\sigma[I_{\ell-1, m} + 1, l]$ but q does not we have $|D(\bar{\omega}, I_{\ell-1, m}(\sigma) + 1, l + 1)| < |D(\omega, I_{\ell-1, m}(\sigma) + 1, l + 1)|$ and (5) holds as well.

To finish the induction, assume we already established $I_{i, m}(\bar{\omega}) \geq I_{i, m}(\omega)$. Then $I_{i+1, m}(\bar{\omega}) \geq I_{i+1, m}(\omega)$ follows by observing that

$$|D(\bar{\omega}, I_{i, m}(\bar{\omega}) + 1, I_{i+1, m}(\omega))| \leq |D(\omega, I_{i, m}(\omega) + 1, I_{i+1, m}(\omega))|,$$

which follows from the fact that $|D(\sigma, i, j)| = |D(\bar{\sigma}^{(p, q)}, i, j)|$ for any sequence σ and $1 \leq i < j \leq |\sigma|$. \square

Remark *i.)* Theorem 3.5 is actually equivalent to Theorem 1 in [4], but we believe that our proof is simpler and more straightforward; it is only half as long, too. *ii.)* Theorem 3.5 also holds for the more general average concave function model [2], as Lemma 1 from [4] holds for this as well. *iii.)* In fact, LRU is optimal for each locality of reference model with the following property: If the sequence $\sigma_1 q \sigma_2 p \sigma_3 q \sigma_4 p \sigma_5$ with sequence σ_3 containing neither p nor q is feasible, then so is the sequence $\sigma_1 q \sigma_2 p \sigma_3 p \sigma_4 q \sigma_5$, which is arguably more “local” than the original one.

Remark Aho et al. [1] study random request sequences that are *almost stationary* in the sense that the page request probabilities maintain their relative orders, i. e., the probability distribution is such that for pages p and q

$$\Pr[p \mid \sigma] \geq \Pr[q \mid \sigma] \implies \Pr[p \mid \sigma\sigma'] \geq \Pr[q \mid \sigma\sigma'] \quad (6)$$

for all request sequences σ and σ' . They define a *ranking relation* $<$ on the pages with $q < p$ meaning $\Pr[q \mid \sigma] \leq \Pr[p \mid \sigma]$ for all request sequences σ . Aho et al. show that the algorithm A_0 which on a page fault evicts a $<$ -minimal page from the cache is optimal w. r. t. to the expected number of page faults. From (6) and Theorem 3.2 it is clear that A_0 is optimal w. r. t. to stochastic dominance of page faults, implying the result from [1]. Note that the almost stationary random sequences include the *independent reference model* [11], in which each requested page is drawn i. i. d. from a fixed distribution.

References

- [1] Alfred V. Aho, Peter J. Denning, and Jeffrey D. Ullman. Principles of optimal page replacement. *J. ACM*, 18(1):80–93, 1971.
- [2] S. Albers, L. M. Favrholt, and O. Giel. On paging with locality of reference. *Journal on Computer and System Sciences*, 70(2):145–175, 2005.
- [3] S. Angelopoulos, R. Dorrigiv, and A. López-Ortiz. On the separation and equivalence of paging strategies. In *Proceedings of the 18th ACM-SIAM symposium on Discrete algorithms*, pages 229–237, 2007.

- [4] S. Angelopoulos and P. Schweitzer. Paging and list update under bijective analysis. In *Proceedings of the 20th ACM-SIAM symposium on Discrete algorithms*, 2009. To appear.
- [5] L. Becchetti. Modeling locality: A probabilistic analysis of LRU and FWF. In *Proceedings of the 12th European Symp. on Algorithms (ESA)*, pages 98–109, 2004.
- [6] L.A. Belady. A study of replacement algorithms for virtual storage computers. *IBM Systems Journal*, 5:78–101, 1966.
- [7] S. Ben-David and A. Borodin. A new measure for the study of online algorithms. *Algorithmica*, 11(1):73–91, 1994.
- [8] A. Borodin and R. El-Yaniv. *Online Computation and Competitive Analysis*. Cambridge University Press, 1998.
- [9] A. Borodin, S. Irani, P. Raghavan, and B. Schieber. Competitive paging with locality of reference. *Journal of Computer and System Sciences*, 50:244–258, 1995.
- [10] J. Boyar, L. M. Favrholdt, and K. S. Larsen. The relative worst-order ratio applied to paging. *Journal on Computer and System Sciences*, 73(6):818–843, 2007.
- [11] Edward G. Coffman, Jr. and Peter J. Denning. *Operating Systems Theory*. Prentice-Hall, 1973.
- [12] P. J. Denning. The working set model of program behavior. *Communications of the ACM*, 11:323–333, 1968.
- [13] P. J. Denning. Working sets past and present. *IEEE Transactions on Software Engineering*, 6:64–84, 1980.
- [14] R. Dorrigiv and A. López-Ortiz. On certain new models for paging with locality of reference. In *WALCOM 2008*, volume 4921 of *Lecture Notes in Computer Science*, pages 200–209, 2008.
- [15] A. Fiat and M. Mendel. Truly online paging with locality of reference. In *FOCS '97: Proceedings of the 38th Annual Symposium on Foundations of Computer Science*, pages 326–335, 1997.
- [16] B. Hiller and T. Vredeveld. Probabilistic analysis of online bin coloring algorithms via stochastic comparison. In *Proceedings of the 16th Annual European Symposium on Algorithms*, Lecture Notes in Computer Science, 2008. to appear.
- [17] Sandy Irani, Anna R. Karlin, and Steven Phillips. Strongly competitive algorithms for paging with locality of reference. *SIAM J. Comput.*, 25(3):477–497, 1996.
- [18] A. Karlin, M. Manasse, L. Rudolph, and D. Sleator. Competitive snoopy paging. *Algorithmica*, 3:79–119, 1988.

- [19] Anna R. Karlin, Steven J. Phillips, and Prabhakar Raghavan. Markov paging. *SIAM J. Comput.*, 30(2):906–922, 2000.
- [20] E. Koutsoupias and C. H. Papadimitriou. Beyond competitive analysis. *SIAM Journal on Computing*, 30(1):300–317, 2000.
- [21] A. Müller and D. Stoyan. *Comparison Models for Stochastic Models and Risks*. John Wiley & Sons, 2002.
- [22] Konstantinos Panagiotou and Alexander Souza. On adequate performance measures for paging. In *STOC '06: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 487–496, 2006.
- [23] D. D. Sleator and R. E. Tarjan. Amortized efficiency of list update and paging rules. *Communications of the ACM*, 28(2):202–208, 1985.
- [24] E. Torng. A unified analysis of paging and caching. *Algorithmica*, 20(2):175–200, 1998.
- [25] N. E. Young. The k -server dual and loose competitiveness for paging. *Algorithmica*, 11(6):525–541, 1994.
- [26] N. E. Young. On-line paging against adversarially biased random inputs. *Journal of Algorithms*, 37(1):218–235, 2000.
- [27] N. E. Young. On-line file caching. *Algorithmica*, 33(3):371–383, 2002.